

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 1, 2001, 15:47:09 ; Search time 210.42 Seconds  
(without alignments)  
10.238 Million cell updates/sec

Title: US-09-331-631A-3\_COPY\_186\_248

Perfect score: 353  
Sequence: 1 KRDPQREYEDCRRRCQQE.....LINPORGSGRVERGEKES 63

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 268485 seqs, 34193795 residues

Total number of hits satisfying chosen parameters: 268485

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

A\_Geneseq\_36:\*

- 1: /SIDSL/gcgdata/geneseq/geneseqp/AA1980.DAT:\*
- 2: /SIDSL/gcgdata/geneseq/geneseqp/AA1981.DAT:\*
- 3: /SIDSL/gcgdata/geneseq/geneseqp/AA1982.DAT:\*
- 4: /SIDSL/gcgdata/geneseq/geneseqp/AA1983.DAT:\*
- 5: /SIDSL/gcgdata/geneseq/geneseqp/AA1984.DAT:\*
- 6: /SIDSL/gcgdata/geneseq/geneseqp/AA1985.DAT:\*
- 7: /SIDSL/gcgdata/geneseq/geneseqp/AA1986.DAT:\*
- 8: /SIDSL/gcgdata/geneseq/geneseqp/AA1987.DAT:\*
- 9: /SIDSL/gcgdata/geneseq/geneseqp/AA1988.DAT:\*
- 10: /SIDSL/gcgdata/geneseq/geneseqp/AA1989.DAT:\*
- 11: /SIDSL/gcgdata/geneseq/geneseqp/AA1990.DAT:\*
- 12: /SIDSL/gcgdata/geneseq/geneseqp/AA1991.DAT:\*
- 13: /SIDSL/gcgdata/geneseq/geneseqp/AA1992.DAT:\*
- 14: /SIDSL/gcgdata/geneseq/geneseqp/AA1993.DAT:\*
- 15: /SIDSL/gcgdata/geneseq/geneseqp/AA1994.DAT:\*
- 16: /SIDSL/gcgdata/geneseq/geneseqp/AA1995.DAT:\*
- 17: /SIDSL/gcgdata/geneseq/geneseqp/AA1996.DAT:\*
- 18: /SIDSL/gcgdata/geneseq/geneseqp/AA1997.DAT:\*
- 19: /SIDSL/gcgdata/geneseq/geneseqp/AA1998.DAT:\*
- 20: /SIDSL/gcgdata/geneseq/geneseqp/AA1999.DAT:\*
- 21: /SIDSL/gcgdata/geneseq/geneseqp/AA2000.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	353	100.0	666	19 W62829	Macadamia integrifolia
2	334	94.6	625	19 W62830	Macadamia integrifolia
3	332	94.1	666	19 W62828	Macadamia integrifolia
4	118.5	33.6	525	19 W62831	Theobroma cacao an
5	118.5	33.6	566	13 R20181	Sequence encoded b
6	118	33.4	590	19 W62832	Gossypium hirsutum
7	103	29.2	637	19 W62837	Hordeum vulgare an
8	74.5	21.1	593	19 W62835	Zea mays antimicro
9	73	20.7	919	10 P93109	human androgen rec
10	73	20.7	919	18 W4783	Androgen receptor
11	73	20.7	919	21 W78914	Human androgen rec
12	72.5	20.5	919	10 P90996	Human androgen rec

13	71.5	20.3	1898	20 Y30795	A human trichohyal
14	71	20.1	154	20 Y33504	Human unliganded a
15	71	20.1	669	19 W37483	Mouse liver cancer
16	71	20.1	918	12 R12223	Human androgen rec
17	71	20.1	918	20 Y33491	Human androgen rec
18	69.5	19.7	611	20 Y29039	T. gondii immunoge
19	69	19.5	28	19 W62841	Stenocarpus sinuat
20	67	19.0	71	20 W09181	Peptide seq ID No:
21	67	19.0	371	20 W73369	Epitope tagged TBP
22	67	19.0	1162	21 Y58500	HHV8 ORF 73 protei
23	66	18.7	436	17 W03662	Human 70K U1 snRNP
24	66	18.7	436	20 Y22342	Human 70K U1 snRNP
25	66	18.7	614	16 R82630	70K autoantigen, p
26	65.5	18.6	35	13 R21079	Antimicrobial maiz
27	65.5	18.6	816	16 R71111	Spinocerebellar at
28	65.5	18.6	816	20 Y33494	Human SCAL protein
29	65.5	18.6	1382	18 W31867	Human metastasis-a
30	65	18.4	1135	21 Y68784	Human ZC1 protein
31	65	18.4	1239	20 Y55931	Amino acid sequenc
32	65	18.4	2074	21 Y54319	Human ZC1 protein
33	65	18.4	2476	20 W67738	Amino acid sequenc
34	64.5	18.3	910	20 W22191	Pig p105 zona pell
35	64	18.1	86	20 W55073	Mouse brain CNG-1
36	64	18.1	86	20 W55078	GST-HD fusion prot
37	64	18.1	1233	20 W55954	Mouse STE20-relate
38	64	18.1	1299	21 Y58633	Protein regulating
39	64	18.1	2023	21 Y54320	Amino acid sequenc
40	63	17.8	112	20 Y04866	Mycobacterium spec
41	63	17.8	126	20 Y04861	Mycobacterium spec
42	63	17.8	262	20 Y29192	Amino acid sequenc
43	63	17.8	365	18 W4971	Chimeric Ewing's s
44	63	17.8	449	19 W47176	Wilms' tumour poly
45	63	17.8	449	21 Y98804	Human WT1 protein

#### ALIGNMENTS

RESULT 1	
W62829	W62829 standard; Protein: 666 AA.
XX	
AC	W62829;
XX	
DT	27-OCT-1998 (first entry)
XX	
DE	Macadamia integrifolia antimicrobial protein.
XX	
KW	antimicrobial protein; infestation; control.
XX	
OS	Macadamia integrifolia.
XX	
EH	Key
FT	Peptide
FT	Location/Qualifiers
FT	1..28
FT	/note="signal peptide"
FT	29..666
FT	/note="mature protein"
XX	
FN	W09827805-A1.
XX	
PD	02-JUL-1998.
XX	
PF	22-DEC-1997; 97MO-AU00874.
XX	
PR	20-DEC-1996; 96AU-0004275.
XX	
PA	(RETR-) COOP RES CENT "TROPICAL PLANT PATHOLOGY.
PI	Bower NI, Gollter KC, Green JL, Manners JM, Marcus JP;
DR	WPI: 1998-377279/32.
DR	N-PSDB: V42311.
XX	

PT Novel anti-microbial protein from e.g. Macadamia integrifolia -  
PT useful for controlling microbial infestations of plants or mammals  
XX  
XX  
PS Claim 1: Page 39-41; 96pp; English.  
XX  
XX  
CC The sequence is that of an antimicrobial protein which can  
CC be used to control microbial infestations in plants and mammalian  
CC animals.  
XX  
SQ Sequence 666 AA;

Query Match 100.0%; Score 353; DB 19; Length 666;  
Best Local Similarity 100.0%; Pred. No. 1.8e-32;  
Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KRDPQOREYEDCRRRCOEOPROQYOCORRCRQORHGSGDLINPORGSGRYEEGEE 60  
DB 186 krdpqreyedcrrrcceqgprqyqqrcreqrqhgdlinpqrgs9ryeegee 245

OY 61 KOS 63  
DB 246 kgs 248

## RESULT 2

W62830  
ID W62830 standard; Protein; 625 AA.

AC W62830;

DT 27-OCT-1998 (first entry)

DE Macadamia integrifolia antimicrobial protein.

XX antimicrobial protein; infestation; control.

OS Macadamia integrifolia.

XX Key Location/Qualifiers

FT Peptide 1..28  
FT /note="signal peptide"

FT Protein 29..666  
FT /note="mature protein"

PN W09827805-A1.

PD 02-JUL-1998.

PF 22-DEC-1997; 97WO-AU00874.

PR 20-DEC-1996; 96AU-0004275.

PA (RETR-) COOP RES CENT TROPICAL PLANT PATHOLOGY.

PI Bower NI, Goulter KC, Green JL, Manners JM, Marcus JP;

DR WPI: 1998-377279/32.

DR N-PSDB; VA2316.

PT Novel anti-microbial protein from e.g. Macadamia integrifolia -  
PT useful for controlling microbial infestations of plants or mammals  
XX  
XX  
PS Claim 1: Page 43-45; 96pp; English.

CC The sequence is that of an antimicrobial protein which can  
CC be used to control microbial infestations in plants and mammalian  
CC animals.  
XX  
SQ Sequence 625 AA;

Query Match 94.6%; Score 334; DB 19; Length 625;

Best Local Similarity 93.7%; Pred. No. 2.5e-30;  
Matches 59; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 KRDPQOREYEDCRRRCOEOPROQYOCORRCRQORHGSGDLINPORGSGRYEEGEE 60  
DB 145 krdpqreyedcrrrcceqgprqyqqrcreqrqhgdlinpqrgs9ryeegee 204

OY 61 KOS 63  
DB 205 kgs 207

## RESULT 3

W62828  
ID W62828 standard; Protein; 666 AA.

AC W62828;

DT 27-OCT-1998 (first entry)

DE Macadamia integrifolia antimicrobial protein.

XX antimicrobial protein; infestation; control.

OS Macadamia integrifolia.

XX Key Location/Qualifiers

FT Peptide 1..28  
FT /note="signal peptide"

FT Protein 29..666  
FT /note="mature protein"

PN W09827805-A1.

PD 02-JUL-1998.

PF 22-DEC-1997; 97WO-AU00874.

PR 20-DEC-1996; 96AU-0004275.

PA (RETR-) COOP RES CENT TROPICAL PLANT PATHOLOGY.

PI Bower NI, Goulter KC, Green JL, Manners JM, Marcus JP;

DR WPI: 1998-377279/32.

DR N-PSDB; VA2310.

PT Novel anti-microbial protein from e.g. Macadamia integrifolia -  
PT useful for controlling microbial infestations of plants or mammals  
XX  
XX  
PS Claim 1: Page 34-36; 96pp; English.

CC The sequence is that of an antimicrobial protein which can  
CC be used to control microbial infestations in plants and mammalian  
CC animals.  
XX  
SQ Sequence 666 AA;

Query Match 94.1%; Score 332; DB 19; Length 666;  
Best Local Similarity 92.1%; Pred. No. 4.5e-30;  
Matches 58; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

OY 1 KRDPQOREYEDCRRRCOEOPROQYOCORRCRQORHGSGDLINPORGSGRYEEGEE 60  
DB 186 krdpqreyedcrrrcceqgprqyqqrcreqrqhgdlinpqrgs9ryeegee 245

OY 61 KOS 63  
DB 246 eqs 248

RESULT 4

```
W62831
ID W62831 standard; Protein; 525 AA.
XX
AC W62831;
XX
DT 27-OCT-1998 (first entry)
XX
DE Theobroma cacao antimicrobial protein.
XX
KM antimicrobial protein; infestation; control.
XX
OS Theobroma cacao.
XX
PN W09827805-A1.
XX
PD 02-JUL-1998.
XX
PF 22-DEC-1997; 97WO-AU00874.
XX
PR 20-DEC-1996; 96AU-0004275.
XX
PA (RETR-) COOP RES CENT TROPICAL PLANT PATHOLOGY.
XX
PI Bower NI, Goulter KC, Green JL, Manners JM, Marcus JP;
XX
DR WPI; 1998-377279/32.
XX
PT Novel anti-microbial protein from e.g. Macadamia integrifolia -
XX
PS useful for controlling microbial infestations of plants or mammals
XX
PS Claim 1; Page 47-49; 96pp; English.
XX
CC The sequence is that of an antimicrobial protein which can
XX
CC be used to control microbial infestations in plants and mammalian
XX
CC animals.
XX
SQ Sequence 525 AA;

Query Match 33.6%; Score 118.5; DB 19; Length 525;
Best Local Similarity 41.2%; Pred. No. 6.3e-06;
Matches 28; Conservative 14; Mismatches 17; Indels 9; Gaps 4;

QY 1 KRDPQREYEDCRRRCEDQ--EPRQYQOCORC---RQQRHGRGDLINPORGSGR 54
   :|||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
DB 35 erdpqr-gyegqqrresateeregeqgeqrcereykeqrqgee--elqrryqgcqgr 91
   :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|

QY 55 YEGGEKQ 62
   :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
DB 92 cdegqgqg 99

RESULT 5
R20181
ID R20181 standard; Protein; 566 AA.
XX
AC R20181;
XX
DT 16-APR-1992 (first entry)
XX
DE Sequence encoded by 67 kD T. cacao protein cDNA.
XX
KM Cocoa; flavour; vlcilin; seed storage protein.
XX
OS Theobroma cacao.
XX
PN W09119801-A.
XX
PD 26-DEC-1991.
XX
PF 07-JUN-1991; 91WO-GB00914.
XX
PR 11-JUN-1990; 90GB-0013016.
```

```
XX
PA (MRSC ) MARS UK LTD.
XX
PI Spencer ME, Hodge R, Deakin EA, Ashton S;
XX
DR WPI; 1992-024418/03.
XX
DR N-PSDB; Q20377.
XX
PT Recombinant cocoa proteins - are responsible for flavour in cocoa
XX
PT beans and produced in large quantities using yeast and bacterial
XX
PT expression vectors
XX
PS Claim 4; Fig 2; 59pp; English.
XX
CC The inventors claim a 67 kD and 31 kD T. cacao protein, and
XX
CC fragments, and encoding DNAs. The 47 kD and 31 kD proteins are
XX
CC derived from the 67 kD precursor. T. cacao protein cDNA was
XX
CC detected in a cDNA library prepared from immature cocoa beans RNA
XX
CC using a probe based on the AA sequence of a CNBr peptide common to
XX
CC the 47 kD and 31 kD polypeptides. Homology searches revealed close
XX
CC homologues between the 67 kD polypeptide and the viciilins, which are
XX
CC seed storage proteins.
XX
SQ Sequence 566 AA;

Query Match 33.6%; Score 118.5; DB 13; Length 566;
Best Local Similarity 41.2%; Pred. No. 6.8e-06;
Matches 28; Conservative 14; Mismatches 17; Indels 9; Gaps 4;

QY 1 KRDPQREYEDCRRRCEDQ--EPRQYQOCORC---RQQRHGRGDLINPORGSGR 54
   :|||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
DB 35 erdpqr-gyegqqrresateeregeqgeqrcereykeqrqgee--elqrryqgcqgr 91
   :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|

QY 55 YEGGEKQ 62
   :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
DB 92 cdegqgqg 99

RESULT 6
W62832
ID W62832 standard; Protein; 590 AA.
XX
AC W62832;
XX
DT 27-OCT-1998 (first entry)
XX
DE Gossypium hirsutum antimicrobial protein.
XX
KM antimicrobial protein; infestation; control.
XX
OS Gossypium hirsutum.
XX
PN W09827805-A1.
XX
PD 02-JUL-1998.
XX
PF 22-DEC-1997; 97WO-AU00874.
XX
PR 20-DEC-1996; 96AU-0004275.
XX
PA (RETR-) COOP RES CENT TROPICAL PLANT PATHOLOGY.
XX
PI Bower NI, Goulter KC, Green JL, Manners JM, Marcus JP;
XX
DR WPI; 1998-377279/32.
XX
PT Novel anti-microbial protein from e.g. Macadamia integrifolia -
XX
PT useful for controlling microbial infestations of plants or mammals
XX
PS Claim 1; Page 49-51; 96pp; English.
XX
CC The sequence is that of an antimicrobial protein which can
```







XX 16-SEP-1999.  
 PD  
 XX  
 PF 11-MAR-1999; 99WO-US05250.  
 XX  
 PR 12-MAR-1998; 98US-0041886.  
 XX  
 PA (BURN-) BURNHAM INST.  
 XX  
 PI Bredesen DE, Rabizadeh S;  
 XX  
 DR WPI; 1999-561617/47.  
 XX  
 PT New proapoptotic dependence peptides, used to develop products for  
 XX treating, e.g. Alzheimer's disease -  
 XX  
 PS Disclosure; Page 178-179; 1999p; English.  
 XX  
 CC This invention describes novel pure proapoptotic dependence peptides  
 CC which comprise a sequence of an active dependence domain selected from  
 CC dependence polypeptides consisting of p75NTR, androgen receptor, DCC,  
 CC huntingtin polypeptide, Machado-Joseph disease gene product, SCA1, SCA2,  
 CC SCA6 and atrophin-1 polypeptide. The proapoptotic peptides are capable  
 CC of inducing cell death and can be used to develop products to mediate or  
 CC inhibit apoptosis. The methods can be used for reducing the severity of  
 CC a proapoptotic dependence domain mediated pathological conditions e.g.  
 CC Huntington's disease, Alzheimer's disease, Kennedy's disease,  
 CC Sphincterobellar ataxia, dentatorubropallidolysian atrophy,  
 CC Machado-Joseph disease, stroke or head trauma. They can also be used for  
 CC reducing the severity of a pathological condition mediated by upregulated  
 CC cell proliferation or cell survival e.g. neoplastic, malignant,  
 CC autoimmune or fibrotic conditions. This sequence represents a human  
 CC unliganded androgen receptor described in the method of the invention.  
 XX  
 SQ Sequence 154 AA:

Query Match 20.1%; Score 71; DB 20; Length 154;  
 Best Local Similarity 32.8%; Pred. No. 0.45;  
 Matches 19; Conservative 12; Mismatches 21; Indels 6; Gaps 1;

OY 5 QOREYEDCRRRCQEQEPKQYQCRCRCRQOHGSGDLINPQRGSGRYEGEEKQ 62  
 :||::| :||::| :||::| :||::| :||::| :||::| :||::| :||::| :||::| :||::|  
 Db 58 qqqqqq-----qqqqqqqqqqqqlsprrqqqqgqspqahrrpplgylvldeeqg 109

RESULT 15  
 W37483  
 ID W37483 standard; Protein; 669 AA.  
 XX  
 AC W37483;  
 XX  
 DT 27-MAR-1998 (first entry)  
 XX  
 DE Mouse liver cancer-originated culture cell growth factor.  
 XX  
 KW Mouse; human; liver cancer-originated culture cell growth factor;  
 KM hhdGF; HET-A; HET-B.  
 XX  
 OS Mus sp.  
 XX  
 JP09313185-A.  
 XX  
 PD 09-DEC-1997.  
 XX  
 PF 27-MAY-1996; 96JP-0131788.  
 XX  
 PR 27-MAY-1996; 96JP-0131788.  
 XX  
 PA (KISH/) KISHIMOTO C.  
 PA (SEKI) SEKISUI CHEM IND CO LTD.  
 XX  
 DR WPI; 1998-080076/08.

DR N-PSDB; V01731.  
 XX  
 PT DNA segment encoding protein homologous to human liver  
 PT cancer-originated culture cell - which may be modified to produce  
 PT polypeptide of at least 5 continuous amino acids, useful in  
 XX producing, e.g. protein HET-A  
 XX  
 PS Claim 10; Page 13-14; 18pp; Japanese.  
 XX

The present sequence represents a protein which has local homology to a  
 CC human liver cancer-originated culture cell isolated from a mouse  
 CC testicle cDNA library. The sequence may (1) be modified to produce a  
 CC polypeptide comprising at least 5 continuous amino acids; (2) a  
 CC polypeptide protein similar to protein of (1), but comprises no protein  
 CC combined naturally and also has 2 amino acid sequences (as given in  
 CC the specification) which may be modified to produce a sequence of at  
 CC least 5 amino acids; (3) a recombinant DNA molecule comprising a vector  
 CC and the DNA segment of (1); (4) a cell comprising (3); (5) producing a  
 CC polypeptide comprising amino acids corresponding to a protein of (1),  
 CC in which a cell comprising (3) is cultured under a conditions enabling  
 CC the expression of (1) to produce the polypeptide which is isolated.  
 CC The above method may be used to produce proteins HET-A and HET-B.  
 XX  
 SQ Sequence 669 AA:

Query Match 20.1%; Score 71; DB 19; Length 669;  
 Best Local Similarity 35.0%; Pred. No. 2;  
 Matches 21; Conservative 8; Mismatches 25; Indels 6; Gaps 1;  
 OY 1 KRDPQOREYEDCRRRCQEQEPKQYQCRCRCRQOHGSGDLINPQRGSGRYEGEEK 60  
 :||::| :||::| :||::| :||::| :||::| :||::| :||::| :||::| :||::| :||::|  
 Db 322 rrdeerrrrlearrrrrrgeee-----lrrlrrgeetkeerrkeraergrssgeeldec 375

Search completed: March 1, 2001, 15:47:10  
 Job time: 235 sec

